FFT-based methods for the mechanics of composites: A general variational framework

Sébastien Brisard, Luc Dormieux

FFT-based methods for the mechanics of composites: a general variational framework

S. Brisard[a,*], L. Dormieux[a]

[a] *Université Paris-Est, UR Navier, Ecole des Ponts ParisTech, 6-8 av. Blaise Pascal, Cité Descartes, Champs-sur-Marne, F-77455 Marne-La-Vallée cedex 2, France*

Abstract

For more than a decade, numerical methods for periodic elasticity, based on the fast Fourier transform, have been used successfully as alternatives to more conventional (fem, bem) numerical techniques for composites. These methods are based on the direct, point-wise, discretization of the Lippmann-Schwinger equation, and a subsequent truncation of underlying Fourier series required for the use of the fast Fourier transform. The basic FFT scheme is very attractive, because of its simplicity of implementation and use. However, it cannot handle pores or rigid inclusions, for which a specific (and significantly more involved) treatment is required. In the present paper, we propose a new FFT-based scheme which is as simple as the basic scheme, while remaining valid for infinite contrasts. Since we adopted an energy principle as an alternative to the Lippmann-Schwinger equation, our scheme is derived within a variational framework. As a by-product, it provides an energetically consistent rule for the homogenization of boundary voxels, a question which has been pending since the introduction of Fourier-based methods.

Key words: Heterogeneous media, Numerical homogenization, Discrete Fourier Transform, Polarization

Introduction

Although of undeniable practical and theoretical use, closed-form estimates of the effective properties of a composite suffer from intrinsic deficiencies which make them unable to capture some fine microstructural effects, such as polydispersivity of inclusions (pores), or local anisotropy. These shortcomings stem from the fact that most estimates are based on the solution to the problem of Eshelby [1], who considered a unique inclusion embedded in an unbounded medium. Eshelby-based estimates can only crudely account for interactions between inclusions (since the inclusion is *isolated* in Eshelby's problem) or size differences between these inclusions (since the embedding medium is unbounded, there is no characteristic length scale the inclusions can be compared to).

Unfortunately, there is no simple alternative to these estimates, and to overcome these shortcomings, one must resort to direct (full-field) simulation of the heterogeneous medium.

*Corresponding author.
Email addresses: sebastien.brisard@enpc.fr (S. Brisard), luc.dormieux@enpc.fr (L. Dormieux)

Among all available numerical techniques, the finite element method is probably the first to come to mind. It has been successfully used for composite materials [2–4], but the complexity of the mesh generation makes its competitiveness questionable. The boundary element method [5] somehow alleviates this burden, since only the interphase boundaries must be discretized.

In the context of highly heterogeneous materials, FFT-based methods are very attractive, for two reasons. First, a full mesh of the microstructure is not required; instead, the calculations are carried out on a regular grid, which allows direct coupling of this numerical tool with 3d imaging techniques (e.g. tomography). Second, Kanit et al. [2], among others, have observed that, compared with kinematic and static uniform boundary conditions, the periodic boundary conditions are known to minimize size-effects for the determination of macroscopic properties.

Periodic boundary conditions naturally lead to Fourier series [6, 7]. Since the seminal paper by Moulinec and Suquet [8], FFT-based methods have been developed by various authors [9–12], and are now almost routinely applied to e.g. rigid-plastic [13] or viscoplastic [14] materials.

In its most basic form, this method is surprisingly simple. It is however not suitable for composites with infinite contrast between the phases. For this class of materials, appropriate extensions of the basic scheme have been proposed [11], but the simplicity of the original scheme is lost.

The existing FFT-methods also raise another question. It is well known (see for example, in the context of the finite element method, [15]) that a satisfactory discretization of the interphase boundary is critical for the estimation of the effective properties. Therefore, when discretizing a real microstructure into voxels, attention should be paid to those voxels which include boundaries. These voxels are effectively composite, and their mechanical properties must be determined accordingly. To the best of our knowledge, no systematic averaging rule has been proposed for these composite voxels.

In this paper, we propose a new version of the FFT-based method for periodic elasticity. By adopting the energy principle of Hashin and Shtrikman [16] as a starting point, instead of a discretized version of the Lippmann-Schwinger equation [17], we come to a slightly different scheme obtained within a variational framework. Although its implementation is very similar to, and indeed as simple as, the basic scheme, its mathematical status is very different. First, it provides a *rigorous* bound on the elastic energy, for *any* refinement of the grid. Second, it is *always* convergent, even in the presence of rigid inclusions or pores, which do not require a specific treatment; besides, it is found to converge faster (by several orders of magnitude) than the basic scheme. Third, it provides an energetically consistent way of homogenizing composite voxels.

In the first part of this paper, we briefly introduce the so-called *basic scheme* [8, 9], in order to emphasize its strengths and weaknesses, and the points from which our new scheme (the so-called *polarization-based scheme*) departs.

The second part of this paper presents the mathematical core of the polarization-based method. We first state the energy principle of Hashin and Shtrikman [16], which is our starting point; all relevant equations are then derived. A *periodized* Green operator is introduced, and indications regarding its numerical evaluation are given. For the sake of simplicity and illustrative purposes, the framework of 2d (plane strain) elasticity will be adopted. The results presented in this paper can readily be generalized to 3d elasticity, albeit at the expense of a slightly increased notational complexity.

This paper closes on two simple examples which illustrate the performance of our new scheme, compared to the basic scheme, and standard finite element calculations.

1. Background

1.1. The elementary problem in periodic homogenization

Let $\Omega \subset \mathbb{R}^2$ be the rectangular unit-cell of an heterogeneous, periodic medium, with local stiffness $\mathbf{c}(\mathbf{x})$. The dimensions of this unit-cell will be denoted $W \times H$, while the basis vectors, parallel to its sides, will be denoted $\mathbf{e}_x$ and $\mathbf{e}_y$. This material is submitted to the macroscopic strain $\mathbf{E}$, and the resulting displacements locally fluctuate about their macroscopic counterpart $\mathbf{E} \cdot \mathbf{x}$. In *periodic* homogenization, the microstructure as well as the fluctuations of the displacement are periodic. Asymptotic expansion techniques [18–20] then show that the solution of the following elementary problem

$$\operatorname{div}[\mathbf{c}(\mathbf{x}) : \boldsymbol{\varepsilon}(\mathbf{x})] = 0 \qquad (\mathbf{x} \in \Omega), \tag{1a}$$

$$2\varepsilon_{ij}(\mathbf{x}) = \partial_i u_j(\mathbf{x}) + \partial_j u_i(\mathbf{x}) \qquad (\mathbf{x} \in \Omega), \tag{1b}$$

$$\mathbf{u}(\mathbf{x} + mW\mathbf{e}_x) = \mathbf{u}(\mathbf{x}) + mW\mathbf{E} \cdot \mathbf{e}_x \qquad (\mathbf{x} \in \mathbb{R}^2, m \in \mathbb{Z}), \tag{1c}$$

$$\mathbf{u}\left(\mathbf{x} + nH\mathbf{e}_y\right) = \mathbf{u}(\mathbf{x}) + nH\mathbf{E} \cdot \mathbf{e}_y \qquad (\mathbf{x} \in \mathbb{R}^2, n \in \mathbb{Z}), \tag{1d}$$

is required in order to determine the macroscopic properties of the periodic medium. In these equations, $\boldsymbol{\varepsilon}(\mathbf{x})$ denotes the microscopic strain deriving from the displacement $\mathbf{u}(\mathbf{x})$; equations (1c) and (1d) express the periodicity of the displacements. The purpose of this paper is to find a numerical approximation of the solution to (1a)–(1d).

1.2. The periodic Lippmann-Schwinger equation

Similarly to non-periodic materials, the introduction of a so-called *reference material* of homogeneous stiffness $\mathbf{c}_0$ can dramatically ease the solution of the problem (1a)–(1d). Indeed, the local constitutive law may be transformed as follows

$$\boldsymbol{\sigma}(\mathbf{x}) = \mathbf{c}_0 : \boldsymbol{\varepsilon}(\mathbf{x}) + [\mathbf{c}(\mathbf{x}) - \mathbf{c}_0] : \boldsymbol{\varepsilon}(\mathbf{x}) = \mathbf{c}_0 : \boldsymbol{\varepsilon}(\mathbf{x}) + \boldsymbol{\tau}(\mathbf{x}), \tag{2}$$

where the *polarization* $\boldsymbol{\tau}(\mathbf{x}) = [\mathbf{c}(\mathbf{x}) - \mathbf{c}_0] : \boldsymbol{\varepsilon}(\mathbf{x})$ has been introduced. Assuming for the time being that $\boldsymbol{\tau}(\mathbf{x})$ is known, it is readily seen that (1a)–(1d) reduces to a simple elasticity problem formulated on a prestressed, homogeneous medium, the solution of which reads

$$\boldsymbol{\varepsilon}(\mathbf{x}) = \mathbf{E} - (\boldsymbol{\Gamma}_0 * \boldsymbol{\tau})(\mathbf{x}),$$

where $\boldsymbol{\Gamma}_0$ is the fourth-rank Green operator for strains [21] associated with the reference material $\mathbf{c}_0$. Recalling that the polarization field is actually not known, the previous equation remains valid, but shall now be considered as an implicit equation in the unknown strain field $\boldsymbol{\varepsilon}(\mathbf{x})$

$$\boldsymbol{\varepsilon}(\mathbf{x}) + \{\boldsymbol{\Gamma}_0 * [(\mathbf{c} - \mathbf{c}_0) : \boldsymbol{\varepsilon}]\}(\mathbf{x}) = \mathbf{E}, \tag{3}$$

which is the well-known Lippmann-Schwinger equation [17]. In periodic elasticity, the convolution product in (3) is most conveniently evaluated in Fourier space

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{k}) = -\hat{\boldsymbol{\Gamma}}_0(\mathbf{k}) : \hat{\boldsymbol{\tau}}(\mathbf{k}) \quad (\mathbf{k} \neq \mathbf{0}), \qquad \hat{\boldsymbol{\varepsilon}}(\mathbf{0}) = \mathbf{E},$$

where the wave-vector $\mathbf{k}$ only takes discrete values

$$\mathbf{k} = \mathbf{k}_{ab} = \frac{2\pi a}{W}\mathbf{e}_x + \frac{2\pi b}{H}\mathbf{e}_y \qquad (a, b \in \mathbb{Z}). \tag{4}$$

When the reference material is isotropic, with shear modulus μ_0 and Poisson ratio ν_0, the Fourier coefficients of the fourth-rank Green operator are known [7] and read, for both two– and three– dimensional elasticity

$$\hat{\mathbf{\Gamma}}_{0,ijhl}(\mathbf{k}) = \frac{1}{4\mu_0}\left(\delta_{ih}n_jn_l + \delta_{il}n_jn_h + \delta_{jh}n_in_l + \delta_{jl}n_in_h\right) - \frac{1}{2\mu_0(1-\nu_0)}n_in_jn_hn_l, \tag{5}$$

where $\mathbf{n} = \mathbf{k}/k$, and $k = |\mathbf{k}|$ ($\mathbf{k} \neq \mathbf{0}$).

1.3. The basic scheme

In this section, we summarize the principles underlying the so-called *basic* scheme for the FFT-based computation of the response of a linear or non-linear composite. This method was first introduced by Moulinec and Suquet [8], who then extended their scheme to non-linear composites [9]. In order to allow for infinite phase contrast Michel et al. [11] further introduced new developments leading to the *augmented Lagrangian* scheme.

Although far from trivial, the basic scheme is beautifully simple. It solves iteratively the Lippmann-Schwinger equation (3), using Neumann series. The iterations read, in the real space

$$\boldsymbol{\varepsilon}^0(\mathbf{x}) = \mathbf{E}, \tag{6a}$$

$$\boldsymbol{\varepsilon}^{n+1}(\mathbf{x}) = \mathbf{E} - \left\{\mathbf{\Gamma}_0 * \left[(\mathbf{c} - \mathbf{c}_0) : \boldsymbol{\varepsilon}^n\right]\right\}(\mathbf{x}). \tag{6b}$$

This scheme draws its efficiency from the fact that in (6b), the direct product $(\mathbf{c} - \mathbf{c}_0) : \boldsymbol{\varepsilon}^n$ is computed in the real space, while the convolution product is computed in the Fourier space. Each iteration of the basic scheme therefore consists in four steps

1. application of the constitutive law in the real space : $\boldsymbol{\tau}^n(\mathbf{x}) = [\mathbf{c}(\mathbf{x}) - \mathbf{c}_0] : \boldsymbol{\varepsilon}^n(\mathbf{x})$,
2. computation of the Fourier coefficients $\hat{\boldsymbol{\tau}}(\mathbf{k}_{ab})$,
3. computation of the convolution product in the Fourier space

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{k}_{ab}) = -\hat{\mathbf{\Gamma}}_0(\mathbf{k}_{ab}) : \hat{\boldsymbol{\tau}}^n(\mathbf{k}_{ab}), \qquad (\mathbf{k}_{ab} \neq \mathbf{0}),$$
$$\hat{\boldsymbol{\varepsilon}}(\mathbf{0}) = \mathbf{E},$$

4. computation of $\boldsymbol{\varepsilon}(\mathbf{x})$ from its Fourier coefficients $\hat{\boldsymbol{\varepsilon}}(\mathbf{k}_{ab})$ by means of a Fourier series.

Of course, these iterations cannot be carried out exactly, and in order to allow for a numerical estimation of the successive iterates $\boldsymbol{\varepsilon}^n$, the unit cell $\Omega = [0; W] \times [0; H]$ must be discretized into $M \times N$ pixels, of size $\frac{W}{M} \times \frac{H}{N}$. Each of the iterates $\boldsymbol{\varepsilon}^n$ is then approximated by a piecewise-constant field, and steps 2. and 4. of the previous scheme are replaced by discrete Fourier transforms, which in turn can be evaluated very efficiently by the fast Fourier transform (FFT). More details on the application of the FFT to this particular problem can be found in [9].

Unfortunately, the price to pay for the simplicity of this scheme is its relatively slow rate of convergence [22], which was established by Michel et al. [11]. These authors even proved that the basic scheme was not convergent in the case of infinite contrast between the phases, which led them to devise the *augmented Lagrangian* scheme.

We should mention at this point the *accelerated scheme* proposed by Eyre and Milton [10] in the context of electrical conductivity, and further extended to elasticity [11, 12]. In this scheme, much faster rates of convergence are achieved by the use of a modified operator, instead of the Green operator for strains $\mathbf{\Gamma}_0$. However, this accelerated scheme is still not convergent when the contrast between the phases is infinite.

2. Fourier-based methods in elasticity: a variational perspective

In the previous section, it has been shown that the two main drawbacks of the basic scheme are its comparatively slow rate of convergence, and inability to provide a solution in the case of an infinite contrast between the phases. Both drawbacks have been overcome, but only at the expense of an increased complexity.

There is another difficulty involved in all existing FFT-based methods, which stems from the use of the discrete Fourier transform to estimate a Fourier series. The problem lies in the fact that a Fourier series has an infinite number of terms, while a discrete Fourier transform is a *finite* sum. When Fourier series are estimated with discrete Fourier transforms, the highest frequencies are therefore omitted, which introduces some discretization errors, as was acknowledged by Moulinec and Suquet [9]. Because of these discretization errors, some of the interesting properties of the continuous basic scheme are lost.

For example, in the continuous version of the basic scheme the iterate $\boldsymbol{\varepsilon}^n(\mathbf{x})$ is at each step kinematically admissible for the problem (1a)–(1d) (since this strain field results from the convolution of a polarization field with the Green operator for strains, see equation (6b)). Hence, the estimated elastic energy at iteration n is an upper-bound of the equilibrium elastic energy. Since in the discrete version of the basic scheme, the convolution product (6b) is only computed approximately, this result is not preserved after discretization.

Also, whether continuous or discrete, the scheme does not indicate how to account for a possible heterogeneity at a sub-pixel scale. When a real microstructure is discretized into a relatively small number of pixels (coarse grid), it is highly probable that each pixel contains more than one phase. What equivalent elastic properties should then this pixel be given?

In this section, we present a slight modification of the basic scheme. While keeping its remarkable simplicity, the resulting scheme is *always* convergent (even for an infinite contrast); besides, numerical experiments tend to prove that it is much faster than the basic scheme. Finally, we show that these properties hold, regardless of the subpixel microstructure, provided that the proposed averaging rule (10) is applied to heterogeneous pixels.

Our scheme uses the same numerical ingredients as the basic scheme of Moulinec and Suquet [9]. Like these authors, we compute direct products in the real space, and convolution products in the reciprocal space; we also adopt piecewise constant approximations of some local field. We do *not* however require the discretized strain field to *solve* the Lippmann-Schwinger equation (3), but to *minimize* the *variational counterpart* of this equation, which is the energy principle of Hashin and Shtrikman [16].

It iterates on approximations of the polarization field, since these are the trial fields for the energy principle used. This new scheme will therefore be refered to as the *polarization-based* scheme.

As the energy principle of Hashin and Shtrikman [16] is the key to our method, and for the sake of completeness, it is first recalled in the next section.

2.1. The energy principle of Hashin and Shtrikman

This principle may be introduced in a very intuitive way by recalling that for *any* choice of the polarization field $\boldsymbol{\tau}(\mathbf{x})$ and the reference material $\mathbf{c}_0$, the strain field $\boldsymbol{\varepsilon}^*(\mathbf{x}) = \mathbf{E} - (\boldsymbol{\Gamma}_0 * \boldsymbol{\tau})(\mathbf{x})$ is kinematically admissible for the problem (1a)–(1d). The principle of Hashin and Shtrikman [16] may therefore be understood as a mere application of the principle of minimum potential energy with $\boldsymbol{\varepsilon}^*$ as a trial field. Provided that the reference medium is *stiffer* ($\mathbf{c}_0 \geq \mathbf{c}(\mathbf{x}), \mathbf{x} \in \Omega$)

than any of the phases in Ω, we get, after simplifications, and for any choice of the polarization field $\boldsymbol{\tau}(\mathbf{x})$, the following upper bound on the elastic energy in the unit cell

$$\frac{1}{2}\overline{\boldsymbol{\varepsilon} : \mathbf{c} : \boldsymbol{\varepsilon}} \leq \frac{1}{2}\mathbf{E} : \mathbf{c}_0 : \mathbf{E} + \overline{\boldsymbol{\tau}} : \mathbf{E} - \frac{1}{2}\overline{\boldsymbol{\tau} : (\mathbf{c} - \mathbf{c}_0)^{-1} : \boldsymbol{\tau}} - \frac{1}{2}\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma}_0 * \boldsymbol{\tau})}, \tag{7}$$

equality being reached when $\boldsymbol{\tau} = (\mathbf{c} - \mathbf{c}_0) : \boldsymbol{\varepsilon}$. (the opposite inequality holds when the reference material is *softer* than any of the phases in Ω). In equation (7), $\boldsymbol{\varepsilon}$ denotes the solution to problem (1a)–(1d), while overlined quantities denote volume average on the unit cell

$$\overline{\mathcal{B}} = \frac{1}{|\Omega|}\int_{\mathbf{x}\in\Omega} \mathcal{B}(\mathbf{x})\, \mathrm{d}^3 x.$$

The variational principle (7) was derived in a very elegant way by Willis [23], who also proved that the quadratic form of $\boldsymbol{\tau}$ on the right hand-side is positive (resp. negative) definite when $\mathbf{c}_0 \geq \mathbf{c}(\mathbf{x})$ (resp $\mathbf{c}_0 \leq \mathbf{c}(\mathbf{x})$) for all $\mathbf{x} \in \Omega$. This is important, because minimization (resp. maximization) of this quadratic form on a finite-dimension subspace of the space of polarization fields will automatically lead to a linear system with a positive (resp. negative) definite matrix.

2.2. The energy principle of Hashin and Shtrikman with pixel-wise constant polarization fields

In this section, we seek an approximate solution to the periodic elasticity problem (1a)–(1d). The approach is very similar to Moulinec and Suquet [8], except for our starting point, which is the Hashin and Shtrikman [16] energy principle, instead of the Lippmann-Schwinger equation. We therefore use the same discretization of the bidimensional unit cell as in section 1.3.

In what follows, greek subscripts denote the indices of the column ($0 \leq \alpha \leq M-1$) and row ($0 \leq \beta \leq N-1$) of pixels. For example, the characteristic function of pixel (α, β) is given by

$$\chi_{\alpha\beta}(\mathbf{x}) = \chi\left[\mathbf{x} - \left(\alpha + \tfrac{1}{2}\right)\tfrac{W}{M}\mathbf{e}_x - \left(\beta + \tfrac{1}{2}\right)\tfrac{H}{N}\mathbf{e}_y\right] \qquad (\mathbf{x} \in \Omega),$$

where we have introduced the generic function $\mathbf{x} \mapsto \chi(\mathbf{x})$

$$\chi\left(x\mathbf{e}_x + y\mathbf{e}_y\right) = \begin{cases} 1 & \text{if } |x| \leq \frac{W}{2M} \text{ and } |y| \leq \frac{H}{2N}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{c}_0$ be the stiffness of a reference medium, which has been chosen and is to remain fixed throughout this section. Keeping in mind that the same line of reasoning would apply (with all inequalities reversed) with a reference medium softer than the softest phase present in Ω, we assume that the reference medium is stiffer than the stiffest phase present in Ω. Therefore, (7) holds without alteration, for any choice of the polarization field $\boldsymbol{\tau}(\mathbf{x})$. With the basic FFT scheme in mind, it is only too natural to write (7) for a pixel-wise constant polarization field

$$\boldsymbol{\tau}(x, y) = \sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1} \chi_{\alpha\beta}(x, y)\, \boldsymbol{\tau}_{\alpha\beta}, \tag{8}$$

where the MN constant tensors $\boldsymbol{\tau}_{\alpha\beta}$ are free parameters. This choice of the polarization field must be substituted into the right-hand side of (7) ; each term is then estimated separately.

First, the evaluation of $\overline{\boldsymbol{\tau}} : \mathbf{E}$ is straightforward, since the volume average of $\boldsymbol{\tau}$ given by (8) simply reads

$$\overline{\boldsymbol{\tau}} = \frac{1}{MN}\sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1} \boldsymbol{\tau}_{\alpha\beta}. \tag{9}$$

Similarly to the basic scheme, the term $\overline{\boldsymbol{\tau} : (\mathbf{c} - \mathbf{c}_0)^{-1} : \boldsymbol{\tau}}$, involving the local constitutive law, is computed in the real space. Introducing the equivalent stiffness $\mathbf{c}_{\alpha\beta}$ defined by

$$\left(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0\right)^{-1} = \frac{MN}{WH} \int_{x,y} \chi_{\alpha\beta}(x,y) \left[\mathbf{c}(x,y) - \mathbf{c}_0\right]^{-1} \mathrm{d}^2 x, \tag{10}$$

it is not difficult to obtain

$$\overline{\boldsymbol{\tau} : (\mathbf{c} - \mathbf{c}_0)^{-1} : \boldsymbol{\tau}} = \frac{1}{MN} \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \boldsymbol{\tau}_{\alpha\beta} : \left(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0\right)^{-1} : \boldsymbol{\tau}_{\alpha\beta}. \tag{11}$$

Incidentally, (10) provides an averaging rule, consistent with our energetic approach, to compute the equivalent stiffness of an heterogeneous pixel. Therefore, one of the questions we raised above is already answered. It should be noted that the determination of appropriate effective properties of composite voxels has already been studied by Toulemonde et al. [15] in the framework of three-dimensional finite elements. Since the numerical method they used to solve the elasticity problem was different, the formula these authors propose differs from (10). However, in both cases, the estimates of the macroscopic properties are dramatically improved when the properties of these boundary pixels (voxels) are carefully selected.

Evaluation of the last term in (7), namely $\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})}$, is more involved, and will therefore be detailed. First, the convolution product $\boldsymbol{\Gamma} * \boldsymbol{\tau}$ is expressed as a Fourier series [7]

$$(\boldsymbol{\Gamma} * \boldsymbol{\tau})(\mathbf{x}) = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{ab}) : \hat{\boldsymbol{\tau}}(\mathbf{k}_{ab}) \exp(\imath \mathbf{k}_{ab} \cdot \mathbf{x}),$$

where $\mathbf{k}_{ab}$ is the discrete wave-vector given by (4), and it is agreed that $\boldsymbol{\Gamma}(\mathbf{0}) = \mathbf{0}$ (otherwise, $(a,b) = (0,0)$ must be excluded from the previous sum). The resulting expression is then contracted with the polarization field $\boldsymbol{\tau}$, and averaged over the unit cell

$$\begin{aligned}
\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})} &= \frac{1}{WH} \int_{\mathbf{x}\in\Omega} \boldsymbol{\tau}(\mathbf{x}) : (\boldsymbol{\Gamma} * \boldsymbol{\tau})(\mathbf{x}) \, \mathrm{d}^2 x \\
&= \frac{1}{WH} \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} \int_{\mathbf{x}\in\Omega} \boldsymbol{\tau}(\mathbf{x}) : \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{ab}) : \hat{\boldsymbol{\tau}}(\mathbf{k}_{ab}) \exp(\imath \mathbf{k}_{ab} \cdot \mathbf{x}) \, \mathrm{d}^2 x.
\end{aligned}$$

The spatial integral in the above expression can be recognized as the complex conjugate (hereafter denoted $\hat{\boldsymbol{\tau}}^*$) of the Fourier coefficient of the polarization field $\boldsymbol{\tau}$ (see Appendix A)

$$\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})} = \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} \hat{\boldsymbol{\tau}}^*(\mathbf{k}_{ab}) : \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{ab}) : \hat{\boldsymbol{\tau}}(\mathbf{k}_{ab}). \tag{12}$$

The Fourier coefficients $\hat{\boldsymbol{\tau}}(\mathbf{k}_{ab})$ of the polarization field must now be evaluated from (8). The calculation is straightforward, and leads to

$$\hat{\boldsymbol{\tau}}(\mathbf{k}_{ab}) = \frac{1}{MN} \operatorname{sinc}\frac{\pi a}{M} \operatorname{sinc}\frac{\pi b}{N} \exp\left[-\imath\pi\left(\frac{a}{M} + \frac{b}{N}\right)\right] \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \exp\left[-2\imath\pi\left(\frac{\alpha a}{M} + \frac{\beta b}{N}\right)\right] \boldsymbol{\tau}_{\alpha\beta}.$$

It is recognized (see Appendix A) that the sum in the above expression is in fact the exact DFT $\{\hat{\boldsymbol{\tau}}_{..}\}$ of the finite data series $\{\boldsymbol{\tau}_{..}\}$

$$\hat{\boldsymbol{\tau}}(\mathbf{k}_{ab}) = \frac{1}{MN} \operatorname{sinc}\frac{\pi a}{M} \operatorname{sinc}\frac{\pi b}{N} \exp\left[-\imath\pi\left(\frac{a}{M} + \frac{b}{N}\right)\right] \hat{\boldsymbol{\tau}}_{ab},$$

which, on subtitution in (12), finally gives

$$\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})} = \frac{1}{(MN)^2} \sum_{a=-\infty}^{+\infty} \sum_{b=-\infty}^{+\infty} \operatorname{sinc}^2 \frac{\pi a}{M} \operatorname{sinc}^2 \frac{\pi b}{N} \hat{\boldsymbol{\tau}}^*_{ab} : \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{ab}) : \hat{\boldsymbol{\tau}}_{ab}.$$

Taking advantage of the fact that the discrete Fourier transform $\hat{\boldsymbol{\tau}}_{ab}$ is $(M, N)-$periodic, the above double sum is rearranged

$$\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})} = \frac{1}{(MN)^2} \sum_{a=0}^{M-1} \sum_{b=0}^{N-1} \hat{\boldsymbol{\tau}}^*_{ab} : \left[\sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \operatorname{sinc}^2 \frac{\pi (a+mM)}{M} \operatorname{sinc}^2 \frac{\pi (b+nN)}{N} \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{a+mM,b+nN}) \right] : \hat{\boldsymbol{\tau}}_{ab}. \quad (13)$$

The above expression can be further simplified by the introduction of the so-called *periodized Green operator*

$$\hat{\boldsymbol{\Gamma}}_{ab} = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \operatorname{sinc}^2 \frac{\pi (a+mM)}{M} \operatorname{sinc}^2 \frac{\pi (b+nN)}{N} \hat{\boldsymbol{\Gamma}}(\mathbf{k}_{a+mM,b+nN}), \quad (14)$$

as well as the sequence of tensors $\{\hat{\boldsymbol{\eta}}_{..}\}$ and their inverse DFT $\{\boldsymbol{\eta}_{..}\}$

$$\hat{\boldsymbol{\eta}}_{ab} = \hat{\boldsymbol{\Gamma}}_{ab} : \hat{\boldsymbol{\tau}}_{ab}, \qquad \boldsymbol{\eta}_{\alpha\beta} = \mathrm{DFT}^{-1}[\hat{\boldsymbol{\eta}}_{..}]_{\alpha\beta}. \quad (15)$$

Substituting these new notations in (13), and applying the Plancherel theorem (see Appendix A), the last term of (7) reduces to the following expression, which is *exact*, in the sense that no series truncation has been performed

$$\overline{\boldsymbol{\tau} : (\boldsymbol{\Gamma} * \boldsymbol{\tau})} = \frac{1}{(MN)^2} \sum_{a=0}^{M-1} \sum_{b=0}^{N-1} \hat{\boldsymbol{\tau}}^*_{ab} : \hat{\boldsymbol{\eta}}_{ab} = \frac{1}{MN} \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \boldsymbol{\tau}^*_{\alpha\beta} : \boldsymbol{\eta}_{\alpha\beta} = \frac{1}{MN} \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \boldsymbol{\tau}_{\alpha\beta} : \boldsymbol{\eta}_{\alpha\beta}, \quad (16)$$

where the last equality results from the fact that the polarization field $\boldsymbol{\tau}_{\alpha\beta}$ is real in the real space. Gathering (9), (11) and (16), we finally obtain a bound on the elastic energy of the unit cell, for any choice of the polarization tensors $\{\boldsymbol{\tau}_{..}\}$

$$\frac{1}{2}\overline{\boldsymbol{\varepsilon} : \mathbf{c} : \boldsymbol{\varepsilon}} \leq \frac{1}{2}\mathbf{E} : \mathbf{c}_0 : \mathbf{E} + \frac{1}{MN} \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \left[\boldsymbol{\tau}_{\alpha\beta} : \mathbf{E} - \frac{1}{2}\boldsymbol{\tau}_{\alpha\beta} : \left(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0\right)^{-1} : \boldsymbol{\tau}_{\alpha\beta} - \frac{1}{2}\boldsymbol{\tau}_{\alpha\beta} : \boldsymbol{\eta}_{\alpha\beta} \right], \quad (17)$$

where $\{\boldsymbol{\eta}_{..}\}$ is given by (15). This bound must now be optimized with respect to the $M \times N$ free parameters $\{\boldsymbol{\tau}_{..}\}$. This is always possible, since Willis [23] has proved that the right-hand side of (7) is a positive definite quadratic form of *any* polarization field $\boldsymbol{\tau}$ (under the assumption that $\mathbf{c}(\mathbf{x}) \leq \mathbf{c}_0$, for all $\mathbf{x} \in \Omega$). In particular, it is a positive definite form on the space of pixel-wise constant polarization fields, where (17) has one unique minimizer.

We propose to regard this minimizer as the $M \times N$ piecewise constant approximation of the *real* polarization field arising in the periodic elasticity problem (2). The associated strain field reads $\boldsymbol{\varepsilon}_{\alpha\beta} = \mathbf{E} - \boldsymbol{\eta}_{\alpha\beta}$. In section 3.1, it will be shown through an example that $\boldsymbol{\varepsilon}_{\alpha\beta}$ provides a good estimate of the true strain field.

2.3. Practical implementation

Optimization of the bound (17) is carried out by solving stationarity conditions. The above discussion prove that these conditions reduce to a non-singular linear system in the unknowns $\{\boldsymbol{\tau}_{..}\}$; this system will be denoted symbolically $A \cdot x = b$, where x denotes the following column-vector

$$x = [\tau_{xx,00}, \tau_{yy,00}, \sqrt{2}\tau_{xy,00}, \tau_{xx,01}, \tau_{yy,01}, \sqrt{2}\tau_{xy,01}, \ldots, \tau_{xx,M-1,N-1}, \tau_{yy,M-1,N-1}, \sqrt{2}\tau_{xy,M-1,N-1}]^T,$$

and b corresponds to the linear part in the quadratic form (17) to be optimized

$$b = [E_{xx}, E_{yy}, \sqrt{2}E_{xy}, \ldots, E_{xx}, E_{yy}, \sqrt{2}E_{xy}]^T.$$

Finally, writing A explicitly would be a very tedious (if not impossible) task, because of the convolution term $\boldsymbol{\tau}_{\alpha\beta} : \boldsymbol{\eta}_{\alpha\beta}$. This strongly suggests to resort to *iterative methods*, which [24]

> "work by repeatedly improving an approximate solution until it is accurate enough. These methods access the coefficient matrix A of the linear system only via the matrix-vector product $y = A \cdot x$ (and perhaps $z = A^T \cdot x$). Thus the user need only supply a subroutine for computing y (and perhaps z) given x, which permits full exploitation of the sparsity or other special structure of A".

Among the numerous methods available, and because of the positive definiteness of the matrix we need to invert, the conjugate gradient method (CG) was selected in this work. According to Barrett et al. [24], all we need to provide the CG method with is a way to compute $A \cdot x$, where A should here be understood as the linear operator underlying the quadratic form (17), and x is formed with all free parameters $\{\boldsymbol{\tau}\}_{..}$. We obviously have

$$A \cdot x = \sum_{\alpha=0}^{M-1} \sum_{\beta=0}^{N-1} \left[\left(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0 \right)^{-1} : \boldsymbol{\tau}_{\alpha\beta} + \boldsymbol{\eta}_{\alpha\beta} \right], \tag{18}$$

where $\{\boldsymbol{\eta}_{..}\}$ is again given by (15). The computation of $A \cdot x$ can therefore be performed in four steps

1. computation of the DFT $\{\hat{\boldsymbol{\tau}}_{..}\}$ of the polarization field $\{\boldsymbol{\tau}_{..}\}$,
2. computation of $\hat{\boldsymbol{\eta}}_{ab} = \hat{\boldsymbol{\Gamma}}_{ab} : \hat{\boldsymbol{\tau}}_{ab}$ in the Fourier space,
3. computation of the inverse DFT $\{\eta_{..}\}$ of the field $\{\hat{\eta}_{..}\}$,
4. computation of $\sum_{\alpha,\beta}[(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0)^{-1} : \boldsymbol{\tau}_{\alpha\beta} + \boldsymbol{\eta}_{\alpha\beta}]$ in the real space.

Of course, the FFT will advantageously be used to compute the required DFTs. The similarity with the basic discrete scheme is striking. It should however again be emphasized that no approximation has been made to arrive at this result.

To conclude this section, we note that the computational cost of one iteration of the CG method is dominated by the evaluation of $A \cdot x$, which requires the same number of FFTs as one iteration of the basic scheme. Therefore, a comparison between the basic and the polarization-based schemes, based of the number of iterations necessary to achieve convergence is fair. Such a comparison will be presented in section 3.1.

2.4. On the periodized Green operator

The introduction of the periodized Green operator $\hat{\boldsymbol{\Gamma}}_{ab}$ defined by (14) follows naturally from the rigorous derivation presented above. This new operator plays the role in the polarization-based scheme of the classical Green operator in the basic scheme. The classical and periodized Green operators are similar in structure, and it is shown in Appendix B that in the limit of infinite resolution ($M, N \to \infty$), both operators coincide. Our numerical experiments show, however, that this convergence is very slow (the relative difference can reach 30 % for $M = N = 1024$); this indicates that introducing the periodized Green operator is the appropriate way of accounting for finite-resolution effects.

Willot and Pellegrini [25] have already observed that the quality of the numerical solution of (1a)–(1d) could be improved if the classical Green operator was replaced by a modified Green operator; the present work confirms this result. To derive their so-called *discrete* Green operator, Willot and Pellegrini [25] first discretize (1a)–(1d) using centered differences; the resulting discrete system is then solved exactly in Fourier space. Our approach departs from this line of reasoning, since the continuous nature of problem (1a)–(1d) is kept, and the resulting minimization problem is optimized in a finite dimension space.

There is no closed-form expression for $\hat{\boldsymbol{\Gamma}}_{ab}$, which must be estimated numerically. Although this is a rather involved task, it needs to be carried out only once for each value of the grid size (2×2, 4×4, 8×8, …, powers of 2 being required by our implementation of the FFT); indeed, the results can be stored and retrieved for later use.

In order to facilitate this numerical estimation, (9) must first be recast in a more convenient form. This is done thanks to the following identity

$$\operatorname{sinc} \frac{\pi\,(a+mM)}{M} = (-1)^m \left[\frac{a+mM}{a}\,\frac{\pi a}{M}\right]^{-1} \sin\frac{\pi a}{M} = (-1)^m\, a\,(a+mM)^{-1} \operatorname{sinc}\frac{\pi a}{M},$$

which does not apply when $a = 0$ and $m = 0$ (in which case the above expression is 1). Assuming first that neither a, nor b are null, we find the following representation of the periodized Green operator

$$\hat{\boldsymbol{\Gamma}}_{ab} = a^2 b^2 \operatorname{sinc}^2 \frac{\pi a}{M} \operatorname{sinc}^2 \frac{\pi b}{N} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} (a+mM)^{-2}\,(b+nN)^{-2}\,\hat{\boldsymbol{\Gamma}}\left(\mathbf{k}_{a+mM,b+nN}\right).$$

If $a = 0$ and $b \neq 0$, only the terms with $m = 0$ remain, and we find

$$\hat{\boldsymbol{\Gamma}}_{0b} = b^2 \operatorname{sinc}^2 \frac{\pi b}{N} \sum_{n=-\infty}^{+\infty} (b+nN)^{-2}\,\hat{\boldsymbol{\Gamma}}\left(\mathbf{k}_{0,b+nN}\right),$$

as well as a similar expression for $b = 0$, $a \neq 0$. When both a and b are null, we need only consider the term $m = 0$, $n = 0$, which gives $\hat{\boldsymbol{\Gamma}}(\mathbf{0})$. Our convention states that this value is zero, and we therefore get

$$\hat{\boldsymbol{\Gamma}}_{00} = \mathbf{0}.$$

Assuming the unit cell to be square ($W = H$) and adopting the same discretization in both x– and y– directions ($M = N$), observation of the general expression of the Green operator (5) shows that $\hat{\boldsymbol{\Gamma}}_{ab}$ can be derived from the components of the following two tensors

$$\mathbf{N}^{\mathrm{II}} = a^2 b^2 \operatorname{sinc}^2 \frac{\pi a}{M} \operatorname{sinc}^2 \frac{\pi b}{N} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} (a+mM)^{-2}\,(b+nN)^{-2}\,\mathbf{n}_{a+mM,b+nN} \otimes \mathbf{n}_{a+mM,b+nN}, \quad (19)$$

and

$$\mathbf{N}^{\mathrm{IV}} = a^2 b^2 \operatorname{sinc}^2 \frac{\pi a}{M} \operatorname{sinc}^2 \frac{\pi b}{N} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} (a+mM)^{-2} (b+nN)^{-2} \, \mathbf{n}_{a+mM,b+nN} \otimes \mathbf{n}_{a+mM,b+nN} \otimes \mathbf{n}_{a+mM,b+nN} \otimes \mathbf{n}_{a+mM,b+nN}, \quad (20)$$

where

$$\mathbf{n}_{ab} = \frac{a\mathbf{e}_x + b\mathbf{e}_y}{\sqrt{a^2+b^2}}.$$

Again, the boundary cases $a = 0$ or $b = 0$ must be treated appropriately. Numerical experiments show that the series for $\mathbf{N}^{\mathrm{II}}$ and $\mathbf{N}^{\mathrm{IV}}$ converge very slowly, making their estimation difficult. The Poisson summation formula [e.g. 26] is always worth trying in such situations, and indeed proved to be quite efficient. As we were not however able to apply this formula to all components of tensors $\mathbf{N}^{\mathrm{II}}$ and $\mathbf{N}^{\mathrm{IV}}$, we devised another estimation method, based on the approximation of integrals by Riemann sums (see Appendix C for a brief description of this numerical scheme in one dimension).

3. Applications

The general optimization problem stated in the previous section, as well as the numerical computation of the periodized Green operator have been implemented in two dimensions (plane strain elasticity). In this section, the properties of the polarization-based scheme are illustrated with two simple applications, shown on figure 1.

For both applications, square grids were used ($M = N$). Besides, the most efficient implementations of the FFT algorithm require the grid-size to be a power of two. Therefore, 8×8, 16×16, $\ldots 128 \times 128$ grids have been considered here.

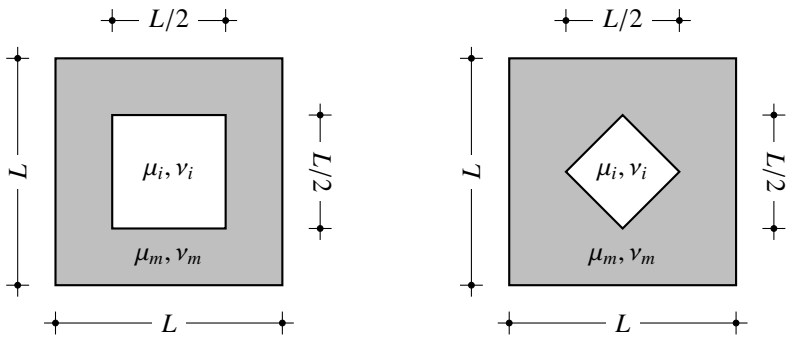

Figure 1: The problem of the square inclusion (left) and the diamond inclusion (right).

3.1. The problem of the square inclusion: comparison with the basic scheme

In this section, we consider a square homogeneous inclusion (shear modulus: μ_i, Poisson ratio: ν_i) embedded in a homogeneous matrix (shear modulus: μ_m, Poisson ratio: ν_m). L being

the size of the (square) unit cell, we set the size of the inclusion to $L/2$ (see figure 1). The unit cell is subjected to an imposed deviatoric macroscopic strain

$$\mathbf{E} = E\left(\mathbf{e}_x \otimes \mathbf{e}_y + \mathbf{e}_y \otimes \mathbf{e}_x\right) \qquad (E = 1),$$

and we want to estimate the average elastic energy in the unit-cell, both with the help of the basic, and the polarization-based schemes. For the application of the basic scheme, the reference material was chosen according to Michel et al. [11]; for the polarization-based scheme, the matrix was chosen as the reference material. Depending on the inclusion being softer or stiffer as the matrix, the resulting estimate of the elastic energy is either an upper- or a lower- bound on the true elastic energy of the composite. Having the reference medium coincide with one of the phases of the true composite requires a special numerical treatment, due to the term $\left(\mathbf{c}_{\alpha\beta} - \mathbf{c}_0\right)$ in (17), which becomes singular. This is accounted for by enforcing that the corresponding local value of the polarization field be zero. This constraint is easily incorporated into a conjugate gradient scheme.

As a first test (figure 2), the contrast between inclusion and matrix is fixed ($\mu_i/\mu_m = 0.1$), and the estimated average elastic energy is plotted against the fineness of the grid (square grids, with $M = N$, were used here). To the best of our knowledge, there is no closed-form solution for the problem at hand. We have therefore superimposed the results of a finite element analysis, carried out with eigh-noded, square elements (based on the same grid), to the FFT-based results.

For numerical schemes providing rigorous bounds on the elasticity, the estimated elastic energy should be a monotonic function of the resolution $M = N = 2^n$. Indeed, the optimum polarization field computed on a $M \times N$ grid by optimization of (17) can be used as a trial field for the refined $2M \times 2N$ grid (since the vertices of the coarser grid match with some vertices of the finer grid). The bound computed on a $M \times N$ grid is therefore always worse than the bound computed on a $2M \times 2N$ grid, and plotting the bound as a function of the resolution leads to a monotonic curve. For the basic scheme, this curve is not monotonic (see figure 2), which shows that the basic scheme does not provide a rigorous bound on the elastic energy; we have already mentioned that this is to be attributed to truncation errors (and the curve becomes indeed much better behaved when the number of Fourier modes N increases). On the other hand, there is no truncation error in our scheme, and we have proved that it provides a bound in all circumstances; this is demonstrated here. It should also be noted that the energy computed from our scheme hardly differs from the finite element estimate.

As a second test, we now turn to the problem of high-contrast, recalling that neither the basic [9], nor the accelerated [10] schemes converge when the contrast μ_i/μ_m tends to zero (pores) or infinity (rigid inclusions), as was proved by Michel et al. [11]. There is no such restriction with the polarization-based scheme, since the matrix to be inverted is always positive (or negative) definite. This matrix might however have turned out to be ill-conditioned for very high contrasts. This was not the case in the examples considered here. Before we comment on figure 3, let us define "convergence" in the present context. In order to ensure fairness of our comparisons, we selected for both basic and polarization-based schemes, similar (though not identical) stopping criteria.

For the basic scheme, first, we simply tested the difference between two successive iterates of the strain field. δ being a small number, the iterations were stopped when

$$\frac{1}{MN}\sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1}\left(\boldsymbol{\varepsilon}_{\alpha\beta}^{n+1} - \boldsymbol{\varepsilon}_{\alpha\beta}^{n}\right) : \left(\boldsymbol{\varepsilon}_{\alpha\beta}^{n+1} - \boldsymbol{\varepsilon}_{\alpha\beta}^{n}\right) \leq \delta\left(\mathbf{E} : \mathbf{E}\right),$$

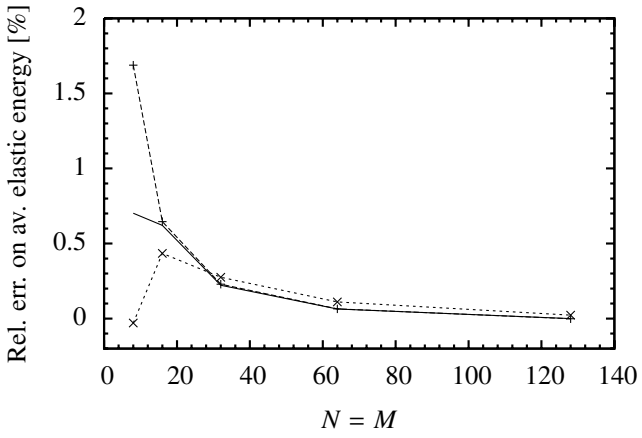

Figure 2: Relative error on the estimated average elastic energy in the problem of the square inclusion, plotted as a function of the resolution ($M = N$ is the number of pixels along each side of the unit-cell). The fem value obtained at the highest resolution is taken as the reference for the definition of the relative error. In this application, $\mu_m = 1$, $\nu_m = 0.3$, $\mu_i = 0.1$ and $\nu_i = 0.2$. The continuous line represents the fem reference calculation, which can hardly be distinguished from the polarization-based results (+). When the resolution is poor, the basic scheme (×) does not provide a bound on the elastic energy.

which is in essence the stopping criterion proposed by Eyre and Milton [10]. Moulinec and Suquet [8, 9] selected a different criterion, based on local equilibrium.

For the polarization-based scheme, we adopted a slighlty more stringent criterion, which was threefold

$$\frac{1}{MN}\sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1}\left(\boldsymbol{\tau}_{\alpha\beta}^{n+1}-\boldsymbol{\tau}_{\alpha\beta}^{n}\right):\left(\boldsymbol{\tau}_{\alpha\beta}^{n+1}-\boldsymbol{\tau}_{\alpha\beta}^{n}\right)\leq\frac{\delta}{MN}\sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1}\boldsymbol{\tau}_{\alpha\beta}^{n}:\boldsymbol{\tau}_{\alpha\beta}^{n},$$

$$\frac{1}{MN}\sum_{\alpha=0}^{M-1}\sum_{\beta=0}^{N-1}\boldsymbol{\rho}_{\alpha\beta}^{n+1}:\boldsymbol{\rho}_{\alpha\beta}^{n+1}\leq\delta\left(\mathbf{E}:\mathbf{E}\right),$$

$$\left|\mathcal{F}^{n+1}-\mathcal{F}^{n}\right|\leq\delta\mathcal{F}^{n},$$

where $\mathcal{F}^n$ denotes the estimated elastic energy and $\boldsymbol{\rho}_{\alpha\beta}^n$ the residual at iteration n, see equations (15) and (18)

$$\boldsymbol{\rho}_{\alpha\beta}^{n}=\left(\mathbf{c}_{\alpha\beta}-\mathbf{c}_0\right)^{-1}:\boldsymbol{\tau}_{\alpha\beta}^{n}+\boldsymbol{\eta}_{\alpha\beta}^{n}-\mathbf{E}.$$

It is recalled that one iteration of the polarization-based scheme is equivalent (in terms of CPU-time) to one iteration of the basic scheme. The comparison presented below is therefore fair. Figure 3 plots, for both basic and polarization-based schemes, the number of iterations necessary to achieve convergence against the contrast μ_i/μ_m. Michel et al. [11] proved that the number of iterations for the basic scheme scales as $\max(\mu_i/\mu_m, \mu_m/\mu_i)$, which is confirmed by this numerical experiment. The asymptotic behaviour of the number of iterations for the polarization-based scheme is much more favourable. Indeed, this number levels off when the contrast is high, so that infinite contrasts can be reached at virtually no additional computational cost. It should further be noticed that for any value of μ_i/μ_m, convergence of the polarization-based scheme is much faster (sometimes by several orders of magnitude) than the basic scheme.

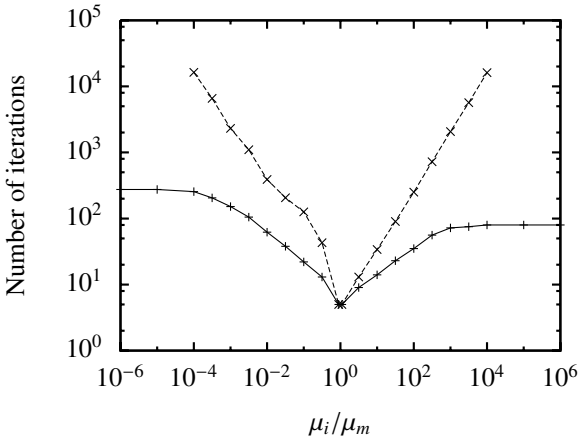

Figure 3: Number of iterations necessary to reach convergence in the problem of the square inclusion ($\mu_m = 1$, $\nu_m = 0.3$, μ_i variable, and $\nu_i = 0.2$). The calculations were carried out on a 128×128 grid, and the stopping criteria described in the text were used with $\delta = 10^{-10}$. The polarization-based scheme (+) always converges in a bounded number of iterations, while the basic scheme ($\times$) requires an increasing number of iterations when the contrast increases.

The two previous tests have shown that the polarization-based scheme converges in a finite number of iterations for any grid-size and elastic contrast. Besides, convergence takes place in a monotonic way, in the sense that the estimated elastic energy is a decreasing function of the resolution. It is then natural to ask how the solution to the minimization problem (17) relates to the solution to the initial problem (1a)–(1d). Since the energy principle of Hashin and Shtrikman states that the minimum in (7) is reached for $\boldsymbol{\tau} = (\mathbf{c} - \mathbf{c}_0) : \boldsymbol{\varepsilon}$, the optimum *discrete* polarization field $\{\boldsymbol{\tau}_{..}\}$ can be seen as an estimate of the optimum *continuous* polarization field. In turn, this discrete polarization field can be used to derive an estimate of the strain field

$$\hat{\boldsymbol{\varepsilon}}_{ab} = -\hat{\boldsymbol{\Gamma}}_{ab} : \hat{\boldsymbol{\tau}}_{ab}, \qquad \boldsymbol{\varepsilon}_{\alpha\beta} = \mathbf{E} + \mathrm{DFT}^{-1}[\hat{\boldsymbol{\varepsilon}}_{..}]_{\alpha\beta}.$$

This calculation was carried out on the problem of the square inclusion, where the estimated local strains derived from the polarization-based method were in very good agreement with their finite element counterpart. This is illustrated on figure 4 in the case of a porous medium ($\mu_i = 0$), for the shear strains ε_{xy} measured along the lower boundary of the unit-cell, see figure 1 (left).

3.2. The problem of the diamond inclusion: discretization errors

The problem we now consider is represented on figure 1 (right). Again, a square homogeneous inclusion (shear modulus: μ_i, Poisson ratio: ν_i) is embedded in a homogeneous matrix (shear modulus: μ_m, Poisson ratio: ν_m), the difference being that the inclusion has been rotated by $\pi/4$ (the inclusion is also smaller). As the matrix–inclusion boundary is oblique, attention must be paid to the boundary pixels, which are composite. To the best of our knowledge, there is no consistent rule to homogenize composite pixels in the basic FFT method. In the polarization-based method, equation (10) provides the averaging rule which must be used in such cases. It could be argued that, asymptotically, any averaging rule would do, since the total area of the boundary pixels tends to zero. This is certainly true of *infinite* resolutions, but not of *finite* resolutions, in which case the status of bound could be lost, as illustrated below.

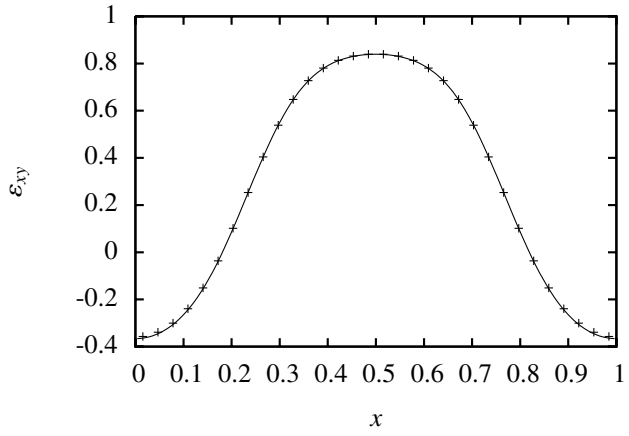

Figure 4: Comparison of the shear strain ε_{xy} obtained with the finite element method (continuous line) and the polarization-based scheme (+), for the problem of the square inclusion. In this example, the inclusion is void ($\mu_i = 0$, $\mu_m = 1$, $\nu_m = 0.3$). The fem solution was computed with 128×128 quadratic elements, while the polarization-based solution was obtained on a 32×32 grid. The strains are measured along the lower boundary of the unit-cell, see figure 1 (left).

Since no comparison with the basic scheme will be carried out in this section, we select the case of an infinite contrast (void inclusion). Two strategies are adopted to carry out the discretization of the problem

- strategy 1: the boundary pixels get their elasticity from equation (10),
- strategy 2: all boundary pixels are arbitrarily affected to the porous phase.

In both strategies, the matrix is chosen as the reference material ($\mu_0 = \mu_m$, $\nu_0 = \nu_m$). The unit-cell is submitted to the same loading as in the previous problem (unit macroscopic shear strain), and a reference calculation is established with the finite element method (six-noded elements). The estimated average elastic energy is represented as a function of the resolution on figure 5. Since the reference medium is stiffer than the matrix, the polarization-based estimate of the elastic energy should be an upper-bound of the true elastic energy. Besides, as the resolution increases, this estimate should dicrease (the minimizer in (17) being seeked on a larger space). This is indeed observed if strategy 1 is applied, but fails to be true with strategy 2, which emphasizes the importance of attributing energetically consistent properties to composite pixels.

It should be noted that even with the correct strategy for the homogenization of boundary pixels, the quality of the upper bound is poorer than in the previous case (compare figure 5 with figure 2). It has been realized, however, that the previous choice of the reference medium ($\mu_0 = \mu_m$, $\nu_0 = \nu_m$) was *not optimal*. For example, selecting $\mu_0 = 17.5$ and $\nu_0 = -1$ leads to a bound which is much closer to the fem estimate.

The question of the choice of the optimum reference medium is usually stated in different terms. Since convergence of the basic iterative scheme is critical, the reference medium is said to be *optimum* if the number of iterations necessary to reach convergence is minimum. We have already shown that convergence is not an issue with the polarization-based method. From the point of view of this method, therefore, the reference medium will be optimum if it leads to the best bound on the elasticity. This question needs to be further investigated.

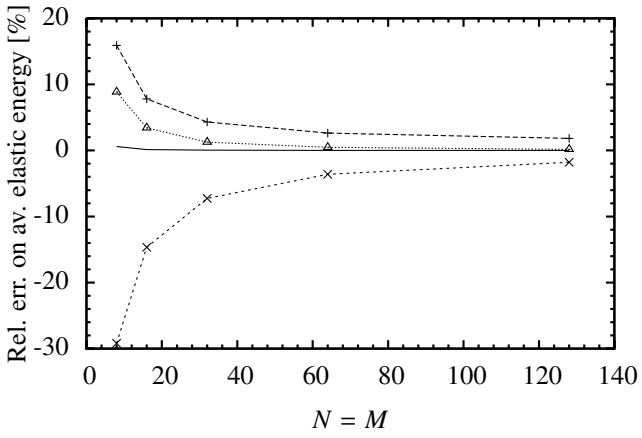

Figure 5: Estimated average elastic energy in the problem of the diamond inclusion, plotted as a function of the resolution. In this application, $\mu_m = 1$, $\nu_m = 0.3$, $\mu_i = 0$. The continuous line represents the fem reference calculation. The polarization-based ($\times$) scheme fails to deliver an upper bound on the energy when strategy 2 is adopted (see text). When equation (10) is applied, this inconsistency disappears (+). By also selecting a better suited reference material, the bound is greatly improved ($\triangle$).

Conclusion

In this paper, we have shown that shifting from the direct discretization of the Lippmann-Schwinger equation to the minimization of the energy of Hashin and Shtrikman [16] on an appropriate subspace shed a new light on FFT-based methods for periodic elasticity.

From the computational point of view, the new, polarization-based scheme is not radically different from the basic scheme. In both methods, a large linear operator needs to be iteratively inverted. At each iteration, updates of the solution are computed by partial applications of this operator in the real and the Fourier space. This clever "operator-splitting" technique was at the origin of the efficiency of the basic scheme [8, 9], the authors of which should be given full credit.

From the theoretical point of view, we believe that the polarization-based scheme improves on the basic scheme. The linear operator underlying this scheme is positive (or negative) definite, which advocates for the use of conjugate gradient methods instead of Neumann series. Convergence is therefore much faster, even at infinite contrast, for which the number of iterations remains bounded. Regardless of the resolution, the new scheme also provides a rigorous bound on the elastic energy. Finally, composite voxels can be accurately accounted for.

Several developments of the polarization-based technique can be thought of. First of all, the multigrid technique proposed by Eyre and Milton [10] should be implemented, in order to try and improve the (already quite high) convergence rate. Second, extensions to nonlinear behaviours, similar to the basic scheme [8, 9] should be considered. Third the determination of the optimum reference medium should be investigated theoretically. Again, it is emphasized that within the framework of the polarization-based scheme, *optimum* should be understood as "leading to the tightest bound on the elastic energy", not as "minimizing the number of iterations to reach convergence".

References

[1] J. D. Eshelby, The determination of the elastic field of an ellipsoidal inclusion, and related problems, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 241 (1226) (1957) 376–396.

[2] T. Kanit, S. Forest, I. Galliet, V. Mounoury, D. Jeulin, Determination of the size of the representative volume element for random composites: statistical and numerical approach, International Journal of Solids and Structures 40 (13–14) (2003) 3647–3679.

[3] S. R. Annapragada, D. Sun, S. V. Garimella, Prediction of effective thermo-mechanical properties of particulate composites, Computational Materials Science 40 (2) (2007) 255–266.

[4] S. Kari, H. Berger, R. Rodriguez-Ramos, U. Gabbert, Computational evaluation of effective material properties of composites reinforced by randomly distributed spherical particles, Composite Structures 77 (2) (2007) 223–231.

[5] Y. Liu, N. Nishimura, Y. Otani, Large-scale modeling of carbon-nanotube composites by a fast multipole boundary element method, Computational Materials Science 34 (2) (2005) 173–187.

[6] S. Nemat-Nasser, T. Iwakuma, M. Hejazi, On composites with periodic structure, Mechanics of Materials 1 (3) (1982) 239–267.

[7] P. Suquet, A simplified method for the prediction of homogenized elastic properties of composites with a periodic structure, Comptes-rendus de l'Académie des sciences série II 311 (7) (1990) 769–774.

[8] H. Moulinec, P. Suquet, A fast numerical method for computing the linear and nonlinear properties of composites, Comptes-rendus de l'Académie des sciences série II 318 (1994) 1417–1423.

[9] H. Moulinec, P. Suquet, A numerical method for computing the overall response of nonlinear composites with complex microstructure, Computer Methods in Applied Mechanics and Engineering 157 (1–2) (1998) 69–94.

[10] D. J. Eyre, G. W. Milton, A fast numerical scheme for computing the response of composites using grid refinement, European Physical Journal-Applied Physics 6 (1) (1999) 41–47, progress in Electromagnetics Research Symposium (PIERS 98), Nantes, France, july 13th–17th, 1998.

[11] J. C. Michel, H. Moulinec, P. Suquet, A computational scheme for linear and non-linear composites with arbitrary phase contrast, International Journal for Numerical Methods in Engineering 52 (1–2) (2001) 139–158, 5th US National Congress on Computational Mechanics, Boulder, co, august 04–06, 1999.

[12] V. Vinogradov, G. W. Milton, An accelerated FFT algorithm for thermoelastic and non-linear composites, Internation Journal for Numerical Methods in Engineering 76 (11) (2008) 1678–1695.

[13] N. Bilger, F. Auslender, M. Bornert, H. Moulinec, A. Zaoui, Bounds and estimates for the effective yield surface of porous media with a uniform or a nonuniform distribution of voids, European Journal of Mechanics - A/Solids 26 (5) (2007) 810 – 836.

[14] M. Idiart, F. Willot, Y.-P. Pellegrini, P. P. Castañeda, Infinite-contrast periodic composites with strongly nonlinear behavior: Effective-medium theory versus full-field simulations, International Journal of Solids and Structures 46 (18-19) (2009) 3365–3382.

[15] C. Toulemonde, R. Masson, J. El Gharib, Modeling the effective elastic behavior of composites: a mixed Finite Element and homogenisation approach, Comptes Rendus Mécanique 336 (3) (2008) 275–282.

[16] Z. Hashin, S. Shtrikman, On some variational principles in anisotropic and nonhomogeneous elasticity, Journal of the Mechanics and Physics of Solids 10 (4) (1962) 335–342.

[17] R. Zeller, P. H. Dederichs, Elastic constants of polycrystals, Physica Status Solidi (b) 55 (2) (1973) 831–842.

[18] A. Bensoussan, J. L. Lions, G. Papanicolaou, Asymptotic analysis for periodic structures, vol. 5 of *Studies in mathematics and its applications*, North-Holland, 1978.

[19] E. Sanchez-Palencia, Comportements local et macroscopique d'un type de milieux physiques hétérogènes, International Journal of Engineering Science 12 (4) (1974) 331–351.

[20] E. Sanchez-Palencia, Non-homogeneous media and vibration theory, vol. 127 of *Lecture Notes in Physics*, Springer-Verlag, 1980.

[21] E. Kröner, Bounds for effective elastic moduli of disordered materials, Journal of the Mechanics and Physics of Solids 25 (2) (1977) 137–155.

[22] H. Moulinec, P. Suquet, Comparison of FFT-based methods for computing the response of composites with highly contrasted mechanical properties, Physica B–Condensed Matter 338 (1–4) (2003) 58–60.

[23] J. R. Willis, Bounds and self-consistent estimates for the overall properties of anisotropic composites, Journal of the Mechanics and Physics of Solids 25 (3) (1977) 185–202.

[24] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, H. V. der Vorst, Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition, SIAM, 1994.

[25] F. Willot, Y.-P. Pellegrini, Fast Fourier transform computations and build-up of plastic deformation in 2D, elastic-perfectly plastic, pixelwise disordered porous media, in: D. Jeulin, S. Forest (Eds.), Continuum models and discrete systems CMDS 11, Presses Mines ParisTech, 443–450, 2008.

[26] A. W. Knapp, Basic real analysis, Cornerstones, Birkhäuser, 2005.

Appendix A. Conventions regarding Fourier series and discrete Fourier transforms

As there is always an ambiguity regarding the signs ($\pm\imath$) and prefactors of Fourier transforms at large, we gather in this section the conventions we adopted throughout this paper. Only one-dimensional Fourier transforms are adressed here, the extension to higher dimensionality being straightforward.

Conventions regarding Fourier series. We consider a univariate function $x \mapsto f(x)$, periodic with period L. We then define the Fourier coefficients of f as

$$\hat{f}(k_a) = \frac{1}{L}\int_0^L f(x)\exp(-\imath k_a x)\,\mathrm{d}\,x,$$

where the wave number k_a takes the discrete values $k_a = 2a\pi/L$. Being periodic, f can be developped as a Fourier series, which reads

$$f(x) = \sum_{a=-\infty}^{+\infty} \hat{f}(k_a)\exp(\imath k_a x).$$

Conventions regarding the discrete Fourier transform (DFT). We now consider a finite sequence of M numbers, $x_0, \ldots, x_{M-1}$. By definition, the DFT of $\{x_\cdot\}$ is the sequence $\{\hat{x}_\cdot\}$

$$\hat{x}_a = \mathrm{DFT}[x_\cdot]_a = \sum_{\alpha=0}^{M-1} x_\alpha \exp\left(-2\imath\pi\frac{\alpha a}{M}\right).$$

Although the above definition holds for any value of $a \in \mathbb{Z}$, it can easily be checked that $\hat{x}_{a+M} = \hat{x}_a$. Therefore, the DFT only defines M different numbers $\hat{x}_0, \ldots, \hat{x}_{M-1}$. In fact, given its DFT, the original sequence may be retrieved through the following inversion formula

$$x_\alpha = \mathrm{DFT}^{-1}[\hat{x}_\cdot]_\alpha = \frac{1}{M}\sum_{a=0}^{M-1} \hat{x}_a \exp\left(2\imath\pi\frac{\alpha a}{M}\right).$$

Finally, we make use in this paper of the Plancherel theorem, which states that the scalar product of two sequences can be performed in the real space or in the Fourier space

$$\sum_{\alpha=0}^{M-1} x_\alpha^* y_\alpha = \frac{1}{M}\sum_{a=0}^{M-1} \hat{x}_a^* \hat{y}_a.$$

Appendix B. Asymptotic behavior of the periodized Green operator

The derivation of the polarization-based scheme requires the introduction of a periodized Green operator, defined by (5). This operator accounts explicitly (and exactly) for the finite resolution ($M, N < \infty$), and a natural question arises: does the periodized Green operator coincide with the original Green operator when the resolution tends to infinity?

The purpose of this section is to show that the answer to this question is yes. Although our proof is not general, practical applications of FFT-based methods generally fall into the case considered below.

In what follows, we restrict ourselves to square unit-cells ($W = H$), with same discretization in both x– and y– directions ($M = N$). Then for fixed values of a and b, we seek the limit

$$\lim_{N\to+\infty} \hat{\mathbf{\Gamma}}_{ab} = \lim_{N\to+\infty} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \operatorname{sinc}^2 \frac{\pi\,(a+mN)}{N} \operatorname{sinc}^2 \frac{\pi\,(b+nN)}{N} \hat{\mathbf{\Gamma}}\Big[(a+mN)\,\mathbf{e}_x + (b+nN)\,\mathbf{e}_y\Big].$$

The above limit can be found by a straightforward application of Lebesgue's dominated convergence theorem. It is therefore necessary to evaluate the limit (when $N \to +\infty$) of each term of the above series. We first have

$$\lim_{N\to+\infty} \operatorname{sinc}^2 \frac{\pi\,(a+mN)}{N} = \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{if } m \neq 0, \end{cases}$$

and it remains then to show that $\hat{\mathbf{\Gamma}}\,(\mathbf{k}_{a+mN,b+nN})$ has a finite limit

$$\lim_{N\to+\infty} \hat{\mathbf{\Gamma}}\,(\mathbf{k}_{a+mN,b+nN}) = \lim_{N\to+\infty} \hat{\mathbf{\Gamma}}\left(\frac{(a+mN)\,\mathbf{e}_x + (b+nN)\,\mathbf{e}_y}{\sqrt{(a+mN)^2 + (b+nN)^2}}\right) = \begin{cases} \hat{\mathbf{\Gamma}}(\mathbf{k}_{ab}) & \text{if } m = 0 \text{ and } n = 0, \\ \hat{\mathbf{\Gamma}}(\mathbf{e}_y) & \text{if } m = 0 \text{ and } n \neq 0, \\ \hat{\mathbf{\Gamma}}(\mathbf{e}_x) & \text{if } m \neq 0 \text{ and } n = 0, \\ \hat{\mathbf{\Gamma}}(m\mathbf{e}_x + n\mathbf{e}_y) & \text{if } m \neq 0 \text{ and } n \neq 0. \end{cases}$$

Gathering the previous equations, we finally find

$$\lim_{N\to+\infty} \hat{\mathbf{\Gamma}}_{ab} = \hat{\mathbf{\Gamma}}(\mathbf{k}_{ab}).$$

Appendix C. On the estimation of some series by way of Riemann sums

Let $f\,(x)$ be a univariate function defined for all $x \geq 0$. We suppose that f is positive, and monotonically decreasing. M being a fixed (positive) number, we want to estimate numerically the following series

$$S\,(a) = \sum_{m=0}^{+\infty} f\,(a+mM)\,.$$

A natural way to do this is to consider the truncated sum

$$S_T\,(a) = \sum_{m=0}^{T} f\,(a+mM)\,,$$

where the truncation value T is chosen large enough to achieve the required accuracy. If the series is slowly convergent, T might be prohibitively large. In order to improve this convergence, let us consider the remainder $R_T\,(a) = S\,(a) - S_T\,(a)$. f being positive and monotically decreasing, the following bounds hold for any m

$$\frac{1}{M}\int_{a+mM}^{a+(m+1)M} f\,(x)\,\mathrm{d}\,x \leq f\,(a+mM) \leq \frac{1}{M}\int_{a+(m-1)M}^{a+mM} f\,(x)\,\mathrm{d}\,x,$$

which, on summation for $m \geq T+1$, gives bounds on the remainder R_T

$$\frac{1}{M} F\left[a + (T+1)\, M\right] \leq R_T\left(a\right) \leq \frac{1}{M} F\left(a + TM\right),$$

where

$$F\left(a\right) = \int_a^{+\infty} f\left(x\right) \mathrm{d}\, x.$$

If $F\left(a\right)$ is known in closed-form, $R_T\left(a\right)$ can be estimated as the mean of the two bounds, the numerical error being bounded by their difference. In other words, the improved estimate of $S\left(a\right)$ reads

$$S_T\left(a\right) + \frac{1}{2M} \left\{F\left[a + (T+1)\, M\right] + F\left(a + TM\right)\right\},$$

and an upper bound on the numerical error is

$$F\left(a + TM\right) - F\left[a + (T+1)\, M\right].$$

Generalization to doubly infinite sums (ranging from $m = -\infty$ to $m = +\infty$), as well as bivariate functions is straightforward. Close inspection of (19) and (20) in fact shows that *i.* all series comply with the assumptions made in this appendix, and *ii.* a closed-form exists for the corresponding required integrals. Numerical experiments showed that the rate of convergence was significantly improved (in the worst case, the error scales as T^{-1} in the initial scheme, and as T^{-2} in our improved scheme).